



UNIVERSITÀ DELLA CALABRIA

DIPARTIMENTO DI ECONOMIA,
STATISTICA E FINANZA
"Giovanni Anania" - DESF

Human Development Index



Relazione finale del corso di “*Modelli Statistici*”

Docente:

Prof. Domma Filippo

Studenti:

Filippelli Marianna

Pulicari Roberta

Ruffo Pietro

Sicilia Mario

CdLM in Statistica & Informatica per le Decisioni e le Analisi di Mercato

A.A. 2019/2020

INDICE

INTRODUZIONE	4
DATA PRE PROCESSING	5
<i>STRUTTURA DEL DATASET</i>	5
INDICE DI SVILUPPO UMANO	10
<i>CALCOLO DELL'ISU</i>	10
<i>ANALISI DESCRITTIVA DELLA VARIABILE RISPOSTA</i>	12
COSTRUZIONE DEL MODELLO	15
<i>TEST SUL PARAMETRO DI PRECISIONE</i>	19
<i>MODELLO FINALE</i>	20
<i>ANALISI DEI RISULTATI</i>	21
CONCLUSIONI E CONSIDERAZIONI DI CARATTERE ECONOMICO	25

INTRODUZIONE

In ogni paese del mondo vi sono persone con prospettive pessimistiche verso il futuro.

Dai margini della società questi individui osservano minoranze elitarie avanzare verso livelli di prosperità via via crescenti. Molti paesi sono riusciti ad emergere dalla povertà estrema tuttavia ancora troppa gente non ha né le opportunità né le risorse per vivere una vita che possa essere definita piena e soddisfacente. Troppo spesso infatti, sesso, etnia o condizioni economiche familiari, determinano, o per meglio dire “condannano”, l’individuo ad una determinata posizione all’interno della società senza prospettiva alcuna di avanzamento.

Le disuguaglianze sono ovunque e in alcune zone del mondo possono rappresentare un affronto alla dignità umana. Le disuguaglianze nello sviluppo umano fanno male alle società, indebolendo la coesione sociale e la fiducia delle persone verso il governo, le istituzioni ed in ultima istanza anche reciproca.

È evidente che lo sviluppo umano non può essere concepito solo nell’ottica economica ma in una prospettiva più ampia che possa abbracciare anche altri ambiti altrettanto importanti per lo sviluppo: è per questo motivo che nel 1990 si avvertì l’esigenza di introdurre l’indice di sviluppo umano.

Questo indice nasce per integrare e superare la visione dello sviluppo come mera crescita economica, espressa dal PIL. L’Indice di Sviluppo Umano è infatti una misura più ampia rispetto al PIL e dà informazioni importanti anche sulla salute e il livello di scolarizzazione di un paese, configurandosi in questo modo come un indice multidimensionale.

L’obiettivo della ricerca è quello di costruire un modello di stima dell’indice di sviluppo umano allo scopo di individuare i fattori che hanno un maggiore impatto sullo stesso nonché metterne in risalto eventuali limiti.

DATA PRE PROCESSING

Struttura del dataset

Il set di dati impiegato nell'analisi è composto da 52 variabili relative a 188 paesi.

Tali variabili sono state ricavate dai seguenti organismi internazionali:

- United Nations Department of Economic and Social Affairs (Dipartimento per gli affari economici e sociali);
- UNESCO Institute for Statistics;
- United Nations Statistics Division;
- Banca mondiale;
- Fondo monetario internazionale.

Allo scopo di una chiara e coerente lettura del dataset si è proceduto ad una classificazione di tipo qualitativo circa l'ampia gamma di variabili a disposizione.

È stato possibile individuare 6 categorie di variabili:

- Economia;
- Sanità e qualità della vita;
- Tecnologia;
- Istruzione;
- Ambiente;
- Società, usi e costumi.

Variabili a carattere economico:

- *Gini coefficient 2005-2013* Il coefficiente di Gini è una misura globale della disuguaglianza nella distribuzione, tra le n unità di una collettività, di un carattere trasferibile (per es. il reddito). È calcolata dopo avere ordinato le n unità secondo l'ammontare non decrescente del carattere posseduto da ciascuna di esse. È spesso usato come indice di concentrazione per misurare la disuguaglianza nella distribuzione del reddito. È un numero compreso tra 0 ed 1. Valori bassi del coefficiente indicano una distribuzione abbastanza omogenea mentre valori alti indicano una distribuzione più diseguale;
- *Total tax revenue of GDP 2005-2013* indica la quota del prodotto interno lordo di un Paese proveniente dalla riscossione delle tasse da parte del governo. Può essere considerato come una misura del grado in cui il governo controlla le risorse dell'economia. Questo indicatore si riferisce

al governo nel suo insieme (tutti i livelli governativi) ed è misurato in milioni di dollari espressi in percentuale del PIL;

- *Consumer price index 2013* è un indice dei prezzi, una misura statistica formata dalla media dei prezzi ponderati per mezzo di uno specifico paniere di beni e servizi. Tale paniere ha come riferimento le abitudini di acquisto di un consumatore medio;
- *Domestic credit provided by financial sector 2013* include tutto il credito concesso a vari settori dell'economia;
- *Domestic food price level 2009 2014 index* è l'indice del prezzo dei prodotti alimentari per uso domestico;
- *Domestic food price level 2009-2014 volatility index* misura la volatilità dei prezzi relativi agli alimenti in un Paese;
- *Exports and imports percentage GDP 2013*: è un indicatore dell'importanza relativa del commercio internazionale nell'economia di un Paese. È calcolato dividendo il valore aggregato delle importazioni e delle esportazioni per un periodo per il prodotto interno lordo per lo stesso periodo;
- *Foreign direct investment net inflows percentage GDP 2013* percentuale di PIL relativa agli investimenti diretti effettuati da investitori non residenti nell'economia del Paese;
- *General government final consumption expenditure - Annual growth 2005 2013* aumento annuale della spesa pubblica per beni e servizi che sono utilizzati per la soddisfazione diretta dei bisogni individuali o dei bisogni collettivi dei membri della comunità;
- *General government final consumption expenditure - % of GDP 2005-2013* percentuale di PIL della spesa pubblica per beni e servizi che sono utilizzati per la soddisfazione diretta dei bisogni individuali o dei bisogni collettivi dei membri della comunità;
- *Gross domestic product GDP 2013* il prodotto interno lordo (PIL) è pari alla somma dei beni e dei servizi finali prodotti da un Paese in un dato periodo di tempo;
- *Gross domestic product GDP per capita* il PIL pro capite è l'indicatore generalmente utilizzato per esprimere il livello di ricchezza per abitante in un territorio in un determinato periodo, consentendo di operare confronti tra aree di dimensione demografica diversa;
- *Gross fixed capital formation of GDP 2005-2013* la formazione lorda di capitale fisso rappresenta i cambiamenti che si verificano in un intervallo di tempo definito per le risorse fisiche legate all'economia di una data nazione. Gli asset fisici coinvolti possono includere miglioramenti apportati agli immobili, la costruzione di strade e altre vie di trasporto, strutture come scuole pubbliche, edifici governativi, fabbriche, ospedali e residenze private;
- *International inbound tourists thousands 2013* numero di turisti stranieri, in migliaia;

- *Private capital flows percentage GDP 2013* i flussi di capitale privato consistono in investimenti diretti in un'impresa che opera in un'economia diversa da quella dell'investitore.
- *Remittances inflows percentual GDP 2013* percentuale di PIL costituita dalle rimesse. La rimessa estera è un trasferimento unilaterale di denaro verso l'estero, effettuato da un lavoratore straniero a beneficio di un altro individuo residente nel suo paese di origine (in genere familiari, parenti o amici).
- *Taxes on income profit and capital gain 2005 2013* le imposte sul reddito, sui profitti e sulle plusvalenze sono riscosse sul reddito netto effettivo o presunto di individui, sugli utili di società e imprese e sulle plusvalenze, realizzate o no, su terreni, titoli e altre attività.

Variabili relative a sanità e qualità della vita:

- *Life expectancy at birth-years* aspettativa di vita all'anno di nascita. È un indicatore statistico che definisce il numero medio di anni di vita di un neonato a tassi di mortalità costanti;
- *Birth registration under age 5 2005-2013* percentuale di registrazioni delle nascite per bambini sotto i 5 anni dal 2005 al 2013. Definisce il numero in percentuale delle nascite effettivamente registrate sotto i primi 5 anni di vita;
- *Female Suicide Rate 100k* tasso di suicidio femminile per 100.000 donne;
- *Homicide rate per 100k people 2008-2012* tasso di omicidio ogni 100.000 persone relativo al periodo 2008-2012;
- *Infant Mortality 2013 per thousands* mortalità infantile relativa all'anno 2013 ogni 1000 bambini. Indica il numero di decessi di bambini nati vivi con età inferiore ad 1 anno.
- *Male Suicide Rate 100k people* tasso di suicidio maschile ogni 100.000 individui;
- *Physicians per 10k people* variabile che indica il numero di dottori, compresi specialisti e generici, in un determinato paese ogni 10 mila residenti;
- *Public health expenditure percentage of GDP 2013* (spesa pubblica investita in sanità, % di PIL ,2013): indica la percentuale di PIL destinata a spese pubbliche sanitarie quali le spese per cure mediche, prevenzione, riabilitazione e tutte le uscite che hanno come obiettivo quello di migliorare la salute;
- *Tuberculosis rate per thousands 2012* indica quante persone risultano malate di tubercolosi su 1000 durante l'anno 2012;
- *Under-five Mortality 2013 thousands* tasso di mortalità bambini sotto i 5 anni ogni mille bambini;

Variabili relative alla tecnologia:

- *Change mobile usage 2009 2014* variazione circa l'utilizzo della connessione mobile nei paesi dal 2009 al 2014;
- *Electrification rate or population* percentuale di popolazione con accesso all'elettricità;

- *Internet users percentage of population 2014* percentuale di utenti internet di una popolazione relativi all'anno 2014;
- *Mobile phone subscriptions per 100 people 2014* abbonamenti di telefonia mobile ogni 100 persone relative all'anno 2014;
- *Research and development expenditure 2005-2012* percentuale di PIL relativa alla spesa in ricerca e sviluppo nel periodo compreso tra il 2005 e il 2012.

Descrizione variabili istruzione:

- *Population with at least some secondary education percent 2005-2013* percentuale di popolazione avente almeno un livello di istruzione secondaria relativa agli anni compresi tra il 2005 e il 2013;
- *International student mobility of total tertiary enrollment 2013* mobilità internazionale studenti sul totale degli iscritti all'università, anno 2013;
- *Primary school dropout rate 2008-2014* tasso di abbandono scolastico di primo grado relativo agli anni tra il 2008 e il 2014;
- *Public expenditure on education Percentage GDP* spesa pubblica investita in istruzione, % di PIL;
- *Pupil-teacher ratio primary school pupils per teacher 2008-2014* numero medio di alunni per insegnante nella scuola elementare e fa riferimento all'arco di tempo compreso tra il 2008 e il 2014;

Descrizione variabili ambientali:

- *Carbon dioxide emissions Average annual growth* incremento annuale medio di emissioni di diossido di carbonio nell'atmosfera;
- *Carbon dioxide emissions per capita 2011 Tones* emissioni di diossido di carbonio pro capite nel 2011 in tonnellate;
- *Change forest percentable 1900 to 2012* percentuale di deforestazione in ciascun paese dal 1900 al 2012;
- *Forest area percentage of total land area 2012* percentuale di area forestale della superficie totale nel 2012;
- *Fossil fuels percentage of total 2012* percentuale di combustibili fossili impiegata per soddisfare il fabbisogno energetico del Paese;
- *Freshwater withdrawals 2005* prelievi di acqua dolce a scopi industriali, agricoli e per usi pubblici e domestici;
- *Homeless people due to natural disaster 2005 2014 per million people* sfollati per calamità naturali per milione di persone tra 2005-2014

- *Natural resource depletion* esaurimento delle risorse naturali in termini di deforestazione, risorse energetiche e minerali;
- *Population affected by natural disasters average annual for million people 2005-2014* media annua di popolazione colpita da catastrofi per milione di persone nel periodo 2005-2014;
- *Population living on degraded land Percentage 2010* percentuale di popolazione che vivono su terreni in stato di degrado. Le stime sul degrado del suolo considerano la biomassa, la salute del suolo, la quantità di acqua e la biodiversità;
- *Renewable sources percentage of total 2012* percentuale di fonti rinnovabili sul consumo totale di energia nel 2012;

Descrizione variabili relative a società, usi e costumi:

- *Intimate or non intimate partner violence ever experienced 2001-2011* percentuale di violenze subite da parte del partner nel periodo 2001-2012;
- *Prison population per 100k people* popolazione carceraria per 100.000 persone;
- *Net migration rate per 1k people 2010-2015* il tasso netto di migrazione per 1.000 abitanti;
- *Refugees by country of origin* percentuale di rifugiati nazionali in altri paesi;
- *Stock of immigrants percentage of population 2013* numero totale di migranti internazionali presenti in un determinato paese nel 2013;
- *Gender Inequality Index 2014* misura le disuguaglianze di genere in tre importanti aspetti dello sviluppo umano: fertilità, emancipazione e la situazione economica.

La fertilità è misurata dal rapporto di mortalità materna e dai tassi di fertilità femminile adolescenziale; l'emancipazione è calcolata dalla percentuale di seggi parlamentari occupati da donne e dalla proporzione di femmine di età pari o superiore a 25 anni con almeno un grado di istruzione secondaria; la situazione economica è espressa tramite la partecipazione delle donne al mercato del lavoro.

INDICE DI SVILUPPO UMANO

L'indice di sviluppo umano, ISU (in inglese: Human Development Index, HDI) è un indicatore di sviluppo macroeconomico realizzato nel 1990 dall'economista pakistano Mahbub ul Haq, seguito inizialmente anche dall'economista indiano Amartya Sen.

L'ISU si affianca al PIL (Prodotto interno lordo), strumento utilizzato convenzionalmente per misurare la ricchezza dei singoli stati, e si iscrive nella logica della misurazione dello sviluppo umano che amplia la prospettiva della semplice crescita economica per definire il livello di sviluppo dei singoli paesi. L'ISU nasce per integrare e superare la visione dello sviluppo come mera crescita economica, espressa dal PIL e dà informazioni importanti anche sulla salute e la scolarizzazione di un paese, configurandosi in questo modo come un indice multidimensionale. Infatti, l'ISU si basa sulla sintesi di tre diversi fattori: il PIL pro capite, l'alfabetizzazione e la speranza di vita.

Calcolo dell'ISU

A partire dal 2010, l'indice di sviluppo umano è definito come la media geometrica di tre indici di base, legati rispettivamente alla speranza di vita, al livello di istruzione e al reddito, i quali a loro volta sono calcolati come precisato di seguito.

L'indice di aspettativa di vita è dato da:

$$\text{Indice di aspettativa di vita (IAV)} = \frac{AV - 20}{85 - 20}$$

Dove AV denota l'aspettativa di vita alla nascita.

Indice di istruzione si calcola come segue:

$$\text{Indice di istruzione (II)} = \frac{IAMI + IAPI}{2}$$

$$\text{Indice anni medi di istruzione (IAMI)} = \frac{AMI}{12}$$

$$\text{Indice anni previsti di istruzione (IAPI)} = \frac{API}{12}$$

Dove AMI rappresenta gli anni medi di istruzione (anni che gli ultraventicinquenni hanno dedicato all'istruzione) e API denota gli anni previsti di istruzione (anni previsti di istruzione per gli infradiciottenni nella loro vita).

Indice di reddito è ottenuto come segue:

$$\text{Indice di reddito (IR)} = \frac{\log(RNLpc) - \log(100)}{\log(75000) - \log(100)}$$

Dove RLNpc denota il reddito nazionale lordo pro capite, corretto in termini di parità di potere d'acquisto ed espresso per convenzione in dollari statunitensi.

L'indice di sviluppo umano è la media geometrica dei tre indici di cui sopra:

$$\text{Indice di sviluppo umano (ISU)} = \sqrt[3]{IAV * II * IR}$$

I valori assunti dall'indice vanno da 0 a 1. Valori prossimi ad 1 corrispondono ad uno sviluppo umano elevato.

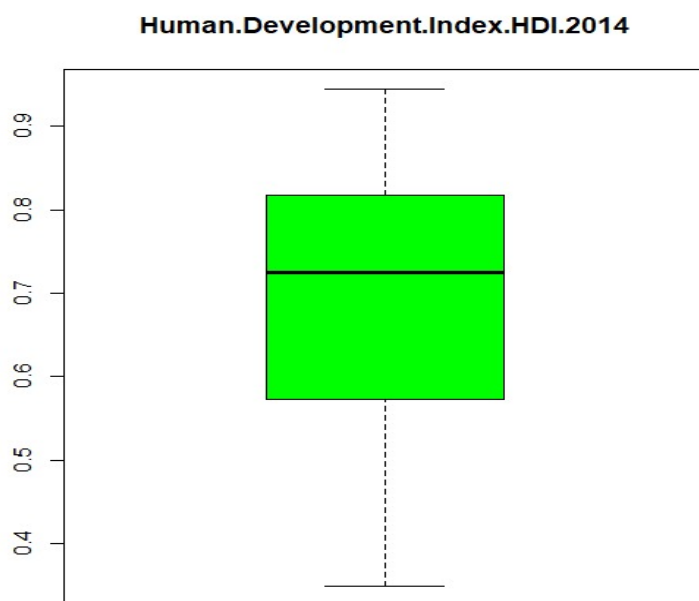
Analisi descrittiva della variabile risposta

Prima di procedere alla costruzione del modello è fondamentale apprezzare la distribuzione della variabile risposta selezionata all'interno del set di dati.

Dal summary è possibile notare che il minimo valore assunto dall'ISU è di 0.3483 corrispondente al Niger mentre il valore massimo coincide con 0.9439 assunto dalla Norvegia.

```
> summary(dati$Human.Development.Index.HDI.2014)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.3483 0.5738 0.7241 0.6924 0.8166 0.9439
```

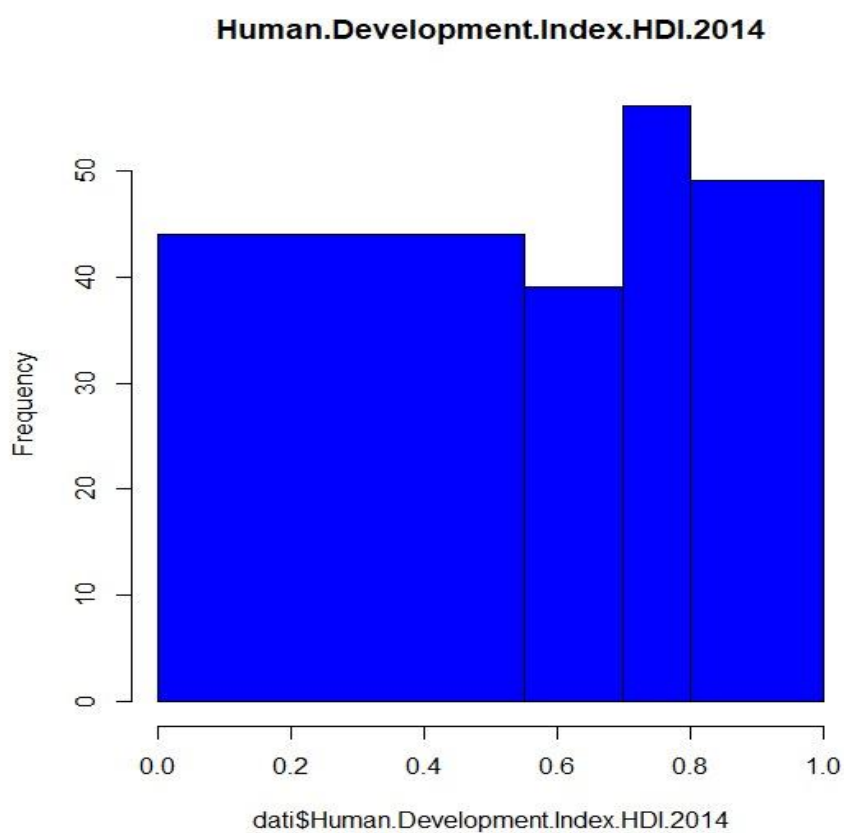
Riportiamo di seguito il boxplot al fine di rappresentare graficamente la variabilità presente nelle osservazioni:



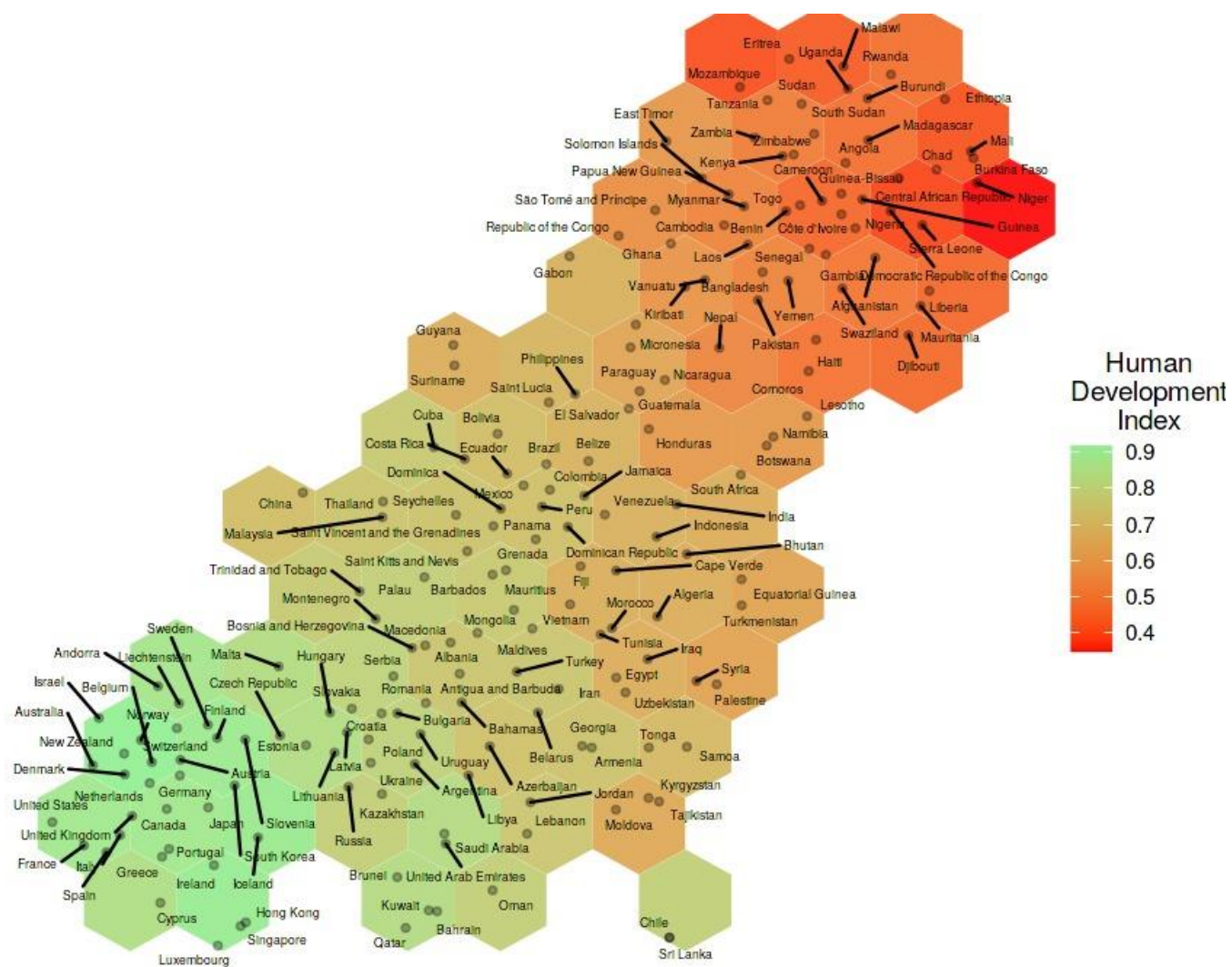
Nonostante il 50 % dei paesi facenti parte dello studio risultano avere un ISU superiore al valore medio (0.6824), vi è un 25 % che presenta valore inferiore a 0.57. Ciò conferma l'eterogeneità dello sviluppo umano nel mondo.

Attraverso l'istogramma si apprezza la frequenza corrispondente alle 4 categorie definite dallo United Nations Department of Economic and Social Affairs:

- Paesi a sviluppo umano molto alto con ISU compreso tra 0.8 e 1;
- Paesi ad alto sviluppo umano con ISU compreso tra 0.7 e 0.799;
- Paesi a medio sviluppo umano con ISU compreso tra 0.550 e 0.699;
- Paesi a basso sviluppo umano con ISU compreso tra 0 e 0.549.



Infine, allo scopo di valutare graficamente la posizione dei paesi in termini di “distanza” e “prossimità” reciproca realtiva al rispettivo valore ISU si riporta il grafico seguente:



Fonte: Kaggle.com

E' semplice notare la notevole distanza, in termini di sviluppo umano che intercorre tra i paesi dell'OCSE occupanti posizioni contraddistinte dal colore verde, e paesi delle aree più povere del mondo come ad esempio l'Africa Subsahariana, contraddistinte dal colore rosso.

COSTRUZIONE DEL MODELLO

Al fine di costruire il modello che possa spiegare al meglio la risposta scelta, si sfrutta la classificazione qualitativa delle variabili effettuata in fase di pre processing per poter effettuare uno screening preliminare di quelle variabili contenute in ciascun gruppo che potenzialmente influenzano il valore dell'Indice di Sviluppo Umano.

Per questo motivo, si applica l'algoritmo di stepwise regression nei vari gruppi, utilizzando come variabile risposta l' ISU e come esplicative, le variabili presenti nei gruppi individuati.

L'algoritmo restituisce il miglior modello lineare costruibile, ossia quel modello che presenta contestualmente al valore AIC più basso, esplicative significative ad un livello di significatività del 15 %.

A scopo esemplificativo, di seguito si riporta il procedimento effettuato per le variabili relative a Salute e Qualità della vita.

```
> variabili_salute<-subset(dati, select=c(Human.Development.Index.HDI.2014,
+ Birth.registration.funder.age.5.2005.2013,
+ Female.Suicide.Rate.100k.people,
+ Homicide.rate.per.100k.people.2008.2012,
+ Infant.Mortality.2013.per.thousands,
+ MaleSuicide.Rate.100k.people,
+ Physicians.per.10k.people,
+ Public.health.expenditure.percentage.of.GDP.2013,
+ Tuberculosis.rate.per.thousands.2012,
+ Under.five.Mortality.2013.thousands))
```

si mostrano le correlazioni elevate utilizzando la funzione “correlazionevariabili()” presente all'interno del file “funzione_correlazione.R”:

```
> source("funzione_correlazione.R")

> (correlazionevariabili(variabili_salute,0.8,F))
               Var1                               Var2      value
50 Under.five.Mortality.2013.thousands Infant.Mortality.2013.per.thousands 0.9914449
```

Si riscontra, prevedibilmente, una correlazione elevatissima tra le variabili Under.five.Mortality.2013.thousands e Infant.Mortality.2013.per.thousands.

Si procede dunque all'eliminazione della variabile Under.five.Mortality.2013.thousands

```
> variabili_salute<-subset(variabili_salute, select=~Under.five.Mortality.2013.thousands)
```


Si applica l'algoritmo di stepwise regression per effettuare lo screening:

```
> #applicazione modello lineare con stepwise regression variabili salute
> modello_variabili_salute_lm<-lm(Human.Development.Index.HDI.2014~.,data=variabili_salute)
> modello_variabili_salute_lm<-stepAIC(modello_variabili_salute_lm, direction = "both")

> summary(modello_variabili_salute_lm)

Call:
lm(formula = Human.Development.Index.HDI.2014 ~ Birth.registration.funder.age.5.2005.2013 +
  Infant.Mortality.2013.per.thousands + Physicians.per.10k.people,
  data = variabili_salute)

Residuals:
    Min       1Q   Median       3Q      Max
-0.21933 -0.03940  0.00278  0.04146  0.22163

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      0.6892490   0.0293980   23.445 < 2e-16 ***
Birth.registration.funder.age.5.2005.2013  0.0006475   0.0002766    2.341  0.0203 *
Infant.Mortality.2013.per.thousands      -0.0040146   0.0003071  -13.073 < 2e-16 ***
Physicians.per.10k.people                 0.0031836   0.0004188    7.602 1.45e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06209 on 184 degrees of freedom
Multiple R-squared:  0.8415,    Adjusted R-squared:  0.8389
F-statistic: 325.6 on 3 and 184 DF,  p-value: < 2.2e-16
```

Una volta terminata la fase di modellazione, funzionale allo screening delle variabili, si utilizza nuovamente l'algoritmo di stepwise regression per costruire un modello lineare eterogeneo ossia contenente simultaneamente tutte le esplicative selezionate nei rispettivi gruppi di appartenenza.

```
> selezione_variabili<-subset(dati,select=c(Human.Development.Index.HDI.2014,
+ # tecnologia
+ Electrification.rate.or.population ,
+ Internet.users.percentage.of.population.2014 ,
+ Mobile.phone.subscriptions.per.100.people.2014 ,
+ Research.and.development.expenditure..2005.2012 ,
+ #salute e qualità della vita
+ Birth.registration.funder.age.5.2005.2013 ,
+ Infant.Mortality.2013.per.thousands ,
+ Physicians.per.10k.people ,
+ #economia
+ Gini.coefficient.2005.2013 ,
+ Domestic.credit.provided.by.financial.sector.2013 ,
+ Domestic.food.price.level.2009.2014.index ,
+ #Ambiente
+ Carbon.dioxide.emissionsAverage.annual.growth ,
+ Carbon.dioxide.emissions.per.capita.2011.Tones ,
+ Change.forest.percentable.1900.to.2012 ,
+ Forest.area.percentage.of.total.land.area.2012 ,
+ Natural.resource.depletion,
+ Population.living.on.degraded.land.Percentage.2010,
+ Renewable.sources.percentage.of.total.2012 ,
+ #Società , usi e costumi
+ Intimate.or.nonintimate.partner.violence.ever.experienced.2001.2011,
+ Prison.population.per.100k.people,Stock.of.immigrants.percentage.of.population.2013,
+ Stock.of.immigrants.percentage.of.population.2013,
+ Gender.Inequality.Index.2014,
+ #istruzione
+ Primary.school.dropout.rate.2008.2014,
+ Pupil.teacher.ratio.primary.school.pupils.per.teacher.2008.2014
+ ))
```

```
> ((correlazionevariabili(selezione_variabili,0.8,F)))
```

	Var1	Var2	value
32	Infant.Mortality.2013.per.thousands	Electrification.rate.or.population	-0.8066029
50	Pupil.teacher.ratio.primary.school.pupils.per.teacher.2008.2014	Electrification.rate.or.population	-0.8277627
73	Gender.Inequality.Index.2014	Internet.users.percentage.of.population.2014	-0.8027604
175	Pupil.teacher.ratio.primary.school.pupils.per.teacher.2008.2014	Infant.Mortality.2013.per.thousands	0.8037749

Dall'analisi di tali correlazioni, quella tra Infant.Mortality.2013.per.thousands ed Electrification.rate.or.population pari a: -0.8066029 porta ad eliminare Infant.Mortality.2013.per.thousands poichè considerata come una chiara conseguenza della seconda variabile.

```
> selezione_variabili<-subset(selezione_variabili,select=~Infant.Mortality.2013.per.thousands)
```

Applicazione dell'algoritmo di stepwise regression sul modello lineare eterogeneo:

```
> #stepwise lm su selezione variabili
> modello_finale_lm<-lm(Human.Development.Index.HDI.2014~.,data=selezione_variabili)
> modello_finale_lm<-stepAIC(modello_finale_lm, direction = "both")

> summary(modello_finale_lm)
```

Call:

```
lm(formula = Human.Development.Index.HDI.2014 ~ Electrification.rate.or.population +
    Internet.users.percentage.of.population.2014 + Mobile.phone.subscriptions.per.100.people.2014 +
    Research.and.development.expenditure..2005.2012 + Physicians.per.10k.people +
    Gini.coefficient.2005.2013 + Population.living.on.degraded.land.Percentage.2010 +
    Renewable.sources.percentage.of.total.2012 + Gender.Inequality.Index.2014 +
    Primary.school.dropout.rate.2008.2014 + Pupil.teacher.ratio.primary.school.pupils.per.teacher.2008.2014,
    data = selezione_variabili)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.106112	-0.026950	0.000331	0.027005	0.100481

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.703e-01	3.551e-02	16.061	< 2e-16 ***
Electrification.rate.or.population	1.023e-03	2.163e-04	4.729	4.60e-06 ***
Internet.users.percentage.of.population.2014	1.803e-03	2.351e-04	7.668	1.14e-12 ***
Mobile.phone.subscriptions.per.100.people.2014	1.997e-04	9.838e-05	2.030	0.04391 *
Research.and.development.expenditure..2005.2012	1.407e-02	5.400e-03	2.605	0.00998 **
Physicians.per.10k.people	5.539e-04	3.311e-04	1.673	0.09617 .
Gini.coefficient.2005.2013	8.409e-04	3.950e-04	2.129	0.03468 *
Population.living.on.degraded.land.Percentage.2010	-4.726e-04	2.647e-04	-1.786	0.07588 .
Renewable.sources.percentage.of.total.2012	-2.280e-04	1.505e-04	-1.515	0.13162
Gender.Inequality.Index.2014	-1.407e-01	3.186e-02	-4.415	1.76e-05 ***
Primary.school.dropout.rate.2008.2014	-1.015e-03	3.360e-04	-3.022	0.00288 **
Pupil.teacher.ratio.primary.school.pupils.per.teacher.2008.2014	-1.152e-03	4.878e-04	-2.361	0.01930 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.04127 on 176 degrees of freedom
Multiple R-squared: 0.933, Adjusted R-squared: 0.9288
F-statistic: 222.9 on 11 and 176 DF, p-value: < 2.2e-16

Data la natura della variabile risposta (continua in (0,1)), la modellazione statistica più adeguata prevede la costruzione di modello di **regressione beta** in cui si utilizzano come base informativa i risultati pervenuti dall'applicazione finale dell'algoritmo di stepwise regression.

```
> library(betareg)
> modello_finale_betareg<-betareg(Human.Development.Index.HDI.2014 ~ Electrification.rate.or.population +
+ Internet.users.percentage.of.population.2014 + Mobile.phone.subscriptions.per.100.people.2014 +
+ Research.and.development.expenditure..2005.2012 + Physicians.per.10k.people +
+ Gini.coefficient.2005.2013 + Population.living.on.degraded.land.Percentage.2010 +
+ Renewable.sources.percentage.of.total.2012 + Gender.Inequality.Index.2014 +
+ Primary.school.dropout.rate.2008.2014 + Pupil.teacher.ratio.primary.school.pupils.per.teacher.2008.2014
+ ,data = selezione_variabili,hessian=T)
> summary(modello_finale_betareg)

Call:
betareg(formula = Human.Development.Index.HDI.2014 ~ Electrification.rate.or.population + Internet.users.percentage.of.population.2014 +
  Mobile.phone.subscriptions.per.100.people.2014 + Research.and.development.expenditure..2005.2012 + Physicians.per.10k.people + Gini.coefficient.2005.2013 +
  Population.living.on.degraded.land.Percentage.2010 + Renewable.sources.percentage.of.total.2012 + Gender.Inequality.Index.2014 + Primary.school.dropout.rate.2008.2014 +
  Pupil.teacher.ratio.primary.school.pupils.per.teacher.2008.2014, data = selezione_variabili, hessian = T)

Standardized weighted residuals 2:
      Min       1Q   Median       3Q      Max
-2.7016 -0.6928 -0.0835  0.7223  2.4520

Coefficients (mean model with logit link):

              Estimate Std. Error z value Pr(>|z|)
(Intercept)      0.3631639   0.1722561    2.108  0.03501 *
Electrification.rate.or.population 0.0032556   0.0010245    3.178  0.00148 **
Internet.users.percentage.of.population.2014 0.0095776   0.0012160    7.877 3.37e-15 ***
Mobile.phone.subscriptions.per.100.people.2014 0.0010499   0.0004957    2.118  0.03418 *
Research.and.development.expenditure..2005.2012 0.1641279   0.0334179    4.911 9.04e-07 ***
Physicians.per.10k.people 0.0037641   0.0018261    2.061  0.03927 *
Gini.coefficient.2005.2013 0.0023458   0.0019570    1.199  0.23066
Population.living.on.degraded.land.Percentage.2010 -0.0022898   0.0012572   -1.821  0.06855 .
Renewable.sources.percentage.of.total.2012 -0.0007206   0.0007600   -0.948  0.34305
Gender.Inequality.Index.2014 -0.7270972   0.1634681   -4.448 8.67e-06 ***
Primary.school.dropout.rate.2008.2014 -0.0041207   0.0015938   -2.585  0.00973 **
Pupil.teacher.ratio.primary.school.pupils.per.teacher.2008.2014 -0.0049085   0.0023114   -2.124  0.03370 *

Phi coefficients (precision model with identity link):
      Estimate Std. Error z value Pr(>|z|)
(phi)  115.7      11.9    9.718 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Type of estimator: ML (maximum likelihood)
Log-likelihood: 342.2 on 13 Df
Pseudo R-squared: 0.9218
Number of iterations in BFGS optimization: 24
```

L'output del modello mette in evidenza la bassa significatività delle covariate Renewable.sources.percentage.of.total.2012 e Gini.coefficient.2005.2013 le quali verranno momentaneamente rimosse dal modello.

```
> modello_finale_betareg<-update(modello_finale_betareg, ~. -Renewable.sources.percentage.of.total.2012-Gini.coefficient.2005.2013)
> summary(modello_finale_betareg)

Call:
betareg(formula = Human.Development.Index.HDI.2014 ~ Electrification.rate.or.population + Internet.users.percentage.of.population.2014 +
  Mobile.phone.subscriptions.per.100.people.2014 + Research.and.development.expenditure..2005.2012 + Physicians.per.10k.people + Population.living.on.degraded.land.Percentage.2010 +
  Gender.Inequality.Index.2014 + Primary.school.dropout.rate.2008.2014 + Pupil.teacher.ratio.primary.school.pupils.per.teacher.2008.2014,
  data = selezione_variabili, hessian = T)

Standardized weighted residuals 2:
      Min       1Q   Median       3Q      Max
-2.7513 -0.7238 -0.0943  0.7507  2.3168

Coefficients (mean model with logit link):

              Estimate Std. Error z value Pr(>|z|)
(Intercept)      0.3855513   0.1529774    2.520 0.011725 *
Electrification.rate.or.population 0.0035571   0.0009889    3.597 0.000322 ***
Internet.users.percentage.of.population.2014 0.0097437   0.0012155    8.016 1.09e-15 ***
Mobile.phone.subscriptions.per.100.people.2014 0.0011401   0.0004943    2.307 0.021076 *
Research.and.development.expenditure..2005.2012 0.1566437   0.0331222    4.729 2.25e-06 ***
Physicians.per.10k.people 0.0036578   0.0018051    2.026 0.042724 *
Population.living.on.degraded.land.Percentage.2010 -0.0021350   0.0012295   -1.737 0.082471 .
Gender.Inequality.Index.2014 -0.7074147   0.1627791   -4.346 1.39e-05 ***
Primary.school.dropout.rate.2008.2014 -0.0041703   0.0015914   -2.621 0.008778 **
Pupil.teacher.ratio.primary.school.pupils.per.teacher.2008.2014 -0.0049072   0.0023197   -2.115 0.034393 *

Phi coefficients (precision model with identity link):
      Estimate Std. Error z value Pr(>|z|)
(phi)  114.31      11.76    9.719 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Type of estimator: ML (maximum likelihood)
Log-likelihood: 341.1 on 11 Df
Pseudo R-squared: 0.9212
Number of iterations in BFGS optimization: 21
```

A questo punto, si vuole verificare se il parametro di precisione è significativamente diverso da 1, cioè:

$$H_0 : \phi = 1 \quad \text{contro} \quad H_1 : \phi \neq 1$$

Statistica test

$$z = \frac{\hat{\phi} - \phi_0}{\sqrt{\text{var}(\hat{\phi})}}$$

```
> z<-((114.31-1)/11.76)
> z
[1] 9.635204
> test1<- 1-(pnorm(z))
> test1
[1] 0
```

Si rifiuta l'ipotesi nulla (H_0) e si procede con la stima del parametro di precisione, sfruttando il test di Breush-Pagan per selezionare i regressori che possano potenzialmente andare a influenzare il parametro in questione. In altre parole, si utilizzano nella stima di ϕ , quei regressori associati ad una elevata significatività nel test BP.

```
> #BP
> residui<-modello_finale_betareg$residuals
> residui<-residui^2
>
> breush_pagan<-lm(residui~Electrification.rate.or.population + Internet.users.percentage.of.population.2014 +
+ Mobile.phone.subscriptions.per.100.people.2014 + Research.and.development.expenditure..2005.2012 +
+ Physicians.per.10k.people + Gender.Inequality.Index.2014 +
+ Primary.school.dropout.rate.2008.2014 +Population.living.on.degraded.land.Percentage.2010+
+ Pupil.teacher.ratio.primary.school.pupils.per.teacher.2008.2014,
+ data=selezione_variabili
+ )
>
> summary(breush_pagan)

Call:
lm(formula = residui ~ Electrification.rate.or.population + Internet.users.percentage.of.population.2014 +
  Mobile.phone.subscriptions.per.100.people.2014 + Research.and.development.expenditure..2005.2012 +
  Physicians.per.10k.people + Gender.Inequality.Index.2014 +
  Primary.school.dropout.rate.2008.2014 + Population.living.on.degraded.land.Percentage.2010 +
  Pupil.teacher.ratio.primary.school.pupils.per.teacher.2008.2014,
  data = selezione_variabili)

Residuals:
    Min       1Q   Median       3Q      Max
-0.0029262 -0.0013070 -0.0005996  0.0006706  0.0132538

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    3.694e-03  1.694e-03   2.180  0.03053 *
Electrification.rate.or.population    6.684e-06  1.114e-05   0.600  0.54925
Internet.users.percentage.of.population.2014 -1.091e-05  1.255e-05  -0.870  0.38571
Mobile.phone.subscriptions.per.100.people.2014 -4.869e-06  5.220e-06  -0.933  0.35220
Research.and.development.expenditure..2005.2012 -1.763e-04  2.845e-04  -0.620  0.53610
Physicians.per.10k.people -5.372e-05  1.753e-05  -3.065  0.00251 **
Gender.Inequality.Index.2014 -3.231e-03  1.684e-03  -1.919  0.05658 .
Primary.school.dropout.rate.2008.2014 -4.038e-05  1.784e-05  -2.263  0.02484 *
Population.living.on.degraded.land.Percentage.2010 1.716e-05  1.383e-05  1.241  0.21627
Pupil.teacher.ratio.primary.school.pupils.per.teacher.2008.2014 4.414e-05  2.607e-05  1.693  0.09220 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.002211 on 178 degrees of freedom
Multiple R-squared:  0.142,    Adjusted R-squared:  0.09862
F-statistic: 3.273 on 9 and 178 DF,  p-value: 0.001035
```

Stima del modello finale in grado non di solo stimare la media della variabile risposta, ma anche il parametro di precisione che ne influenza la varianza:

```
> summary(mf_con_phi)

Call:
betareg(formula = Human.Development.Index.HDI.2014 ~ Electrification.rate.or.population + Internet.users.percentage.of.population.2014 +
  Research.and.development.expenditure..2005.2012 + Physicians.per.10k.people + Gender.Inequality.Index.2014 + Primary.school.dropout.rate.2008.2014 +
  Pupil.teacher.ratio.primary.school.pupils.per.teacher.2008.2014 | Physicians.per.10k.people + Primary.school.dropout.rate.2008.2014 +
  Gender.Inequality.Index.2014 + Pupil.teacher.ratio.primary.school.pupils.per.teacher.2008.2014, data = selezione_variabili,
  hessian = T)

Standardized weighted residuals 2:
      Min       IQ   Median       3Q        Max
-2.4632 -0.7407 -0.0506  0.7417  2.9668

Coefficients (mean model with logit link):
                                Estimate Std. Error z value Pr(>|z|)
(Intercept)                    0.4238773   0.1188874   3.565 0.000363 ***
Electrification.rate.or.population 0.0047106   0.0008908   5.288 1.24e-07 ***
Internet.users.percentage.of.population.2014 0.0097062   0.0008372  11.594 < 2e-16 ***
Research.and.development.expenditure..2005.2012 0.1399893   0.0328244   4.265 2.00e-05 ***
Physicians.per.10k.people        0.0044297   0.0008006   5.533 3.15e-08 ***
Gender.Inequality.Index.2014     -0.9068558   0.1149752  -7.887 3.09e-15 ***
Primary.school.dropout.rate.2008.2014 -0.0022923   0.0013189  -1.738 0.082194 .
Pupil.teacher.ratio.primary.school.pupils.per.teacher.2008.2014 -0.0045777   0.0020544  -2.228 0.025863 *

Phi coefficients (precision model with log link):
                                Estimate Std. Error z value Pr(>|z|)
(Intercept)                    2.64076   0.65512   4.031 5.55e-05 ***
Physicians.per.10k.people        0.06158   0.01525   4.039 5.36e-05 ***
Primary.school.dropout.rate.2008.2014 0.04131   0.01202   3.436 0.000590 ***
Gender.Inequality.Index.2014     3.68814   1.07708   3.424 0.000617 ***
Pupil.teacher.ratio.primary.school.pupils.per.teacher.2008.2014 -0.03259   0.01720  -1.895 0.058142 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Type of estimator: ML (maximum likelihood)
Log-likelihood: 350 on 13 Df
Pseudo R-squared: 0.9178
Number of iterations in BFGS optimization: 26
```

Poiché lo pseudo R^2 assume valori compresi tra 0 e 1, e una perfetta concordanza tra il predittore lineare e le ordinate stimate corrisponde ad uno pseudo $R^2=1$ ¹, si conclude che il modello stimato, avente pseudo $R^2=0.9178$ riesce ad approssimare adeguatamente i dati osservati.

Inoltre il confronto tra i valori assunti dall' Akaike Information Criterion (AIC) nei modelli che presentano rispettivamente il solo predittore lineare per la media, ed il predittore sia per la media che per il parametro di precisione, conferma che è da preferire il modello avente entrambi i predittori.

```
> #confronto aic tra modello_finale_betareg che ha solo il predittore lineare per la media e
> #mf_con_phi che presenta anche il predittore per il parametro di precisione
> AIC(modello_finale_betareg)
[1] -660.1631
> AIC(mf_con_phi)
[1] -674.0745
```

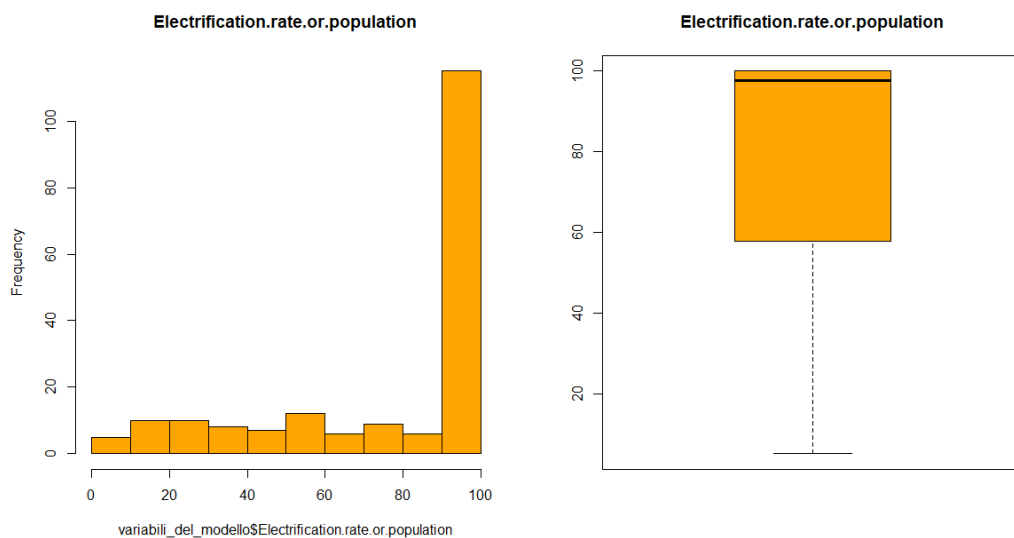
¹ "BETA REGRESSION FOR MODELLING RATES AND PROPORTIONS" SILVIA L.P. FERRARI, FRANCISCO CRIBARI-NETO

Occorre a questo punto esaminare le variabili risultate significative nello spiegare l'ISU alla fine del processo di modellazione.

- *Electrification.rate.or.population*

L'accesso all'energia elettrica, nel 21 esimo secolo, risulta fondamentale per lo sviluppo sociale ed economico della nazione. All'interno del set di dati è possibile rilevare che:

```
> summary(variabili_del_modello$Electrification.rate.or.population)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   5.10   58.60   97.70   77.92  100.00  100.00
```

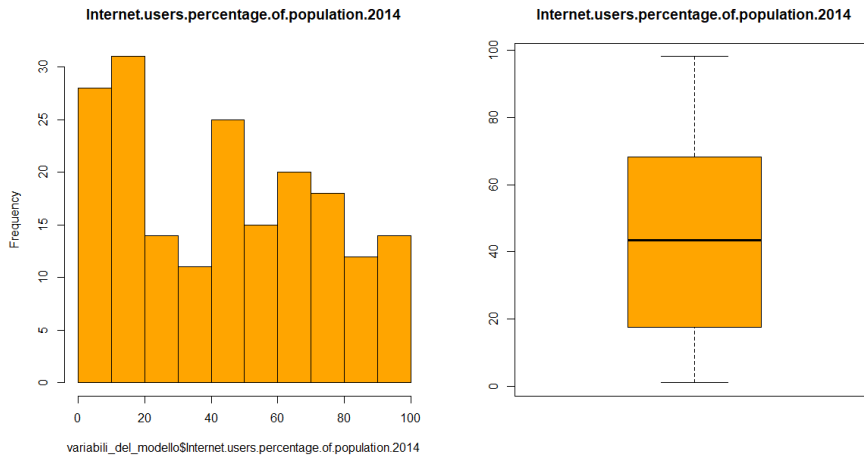


il 25% dei paesi contenuti all'interno del set risultano avere una percentuale di copertura elettrica inferiore al 58.6 %.

- *Internet.users.percentage.of.population.2014*

La percentuale di utenti in rete all'interno di un paese è indicatore del grado di libertà di espressione e conseguentemente di apertura culturale ed economica verso l'esterno.

```
> summary(variabili_del_modello$Internet.users.percentage.of.population.2014)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.99  17.57   43.40   44.11  68.19   98.16
```

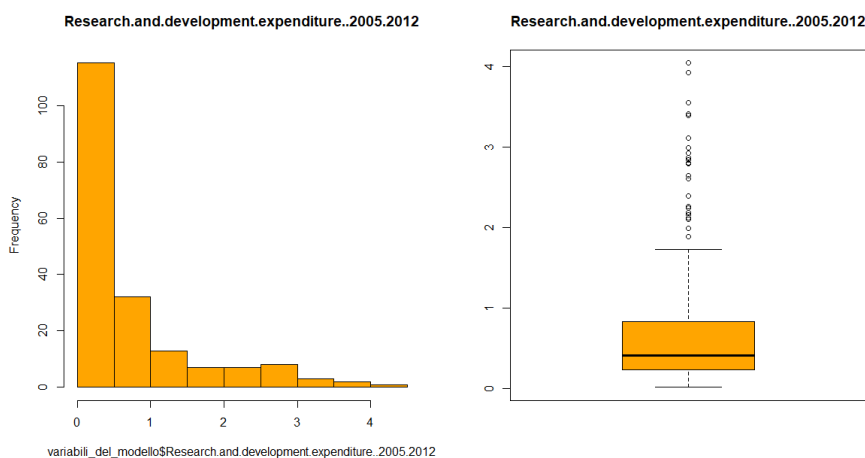


E' evidente che il 50% dei paesi analizzati presenta una percentuale di popolazione con connessione internet inferiore al 43.40%.

- *Research.and.development.expenditure..2005.2012*

La percentuale di PIL investita in ricerca e sviluppo è sinonimo del grado di avanzamento dell'economia di un paese e indirettamente anche del suo livello di sviluppo umano.

```
> summary(variabili_del_modello$Research.and.development.expenditure..2005.2012)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.01283 0.22754 0.41293 0.74327 0.82796 4.03919
```



Solo il 25% dei paesi analizzati, investe più dello 0.8 % del suo PIL in R&D.

- *Gender.Inequality.Index.2014 (GII)*

```
> summary(variabili_del_modello$Gender.Inequality.Index.2014)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.01642 0.20843 0.38669 0.36448 0.50354 0.74396
```

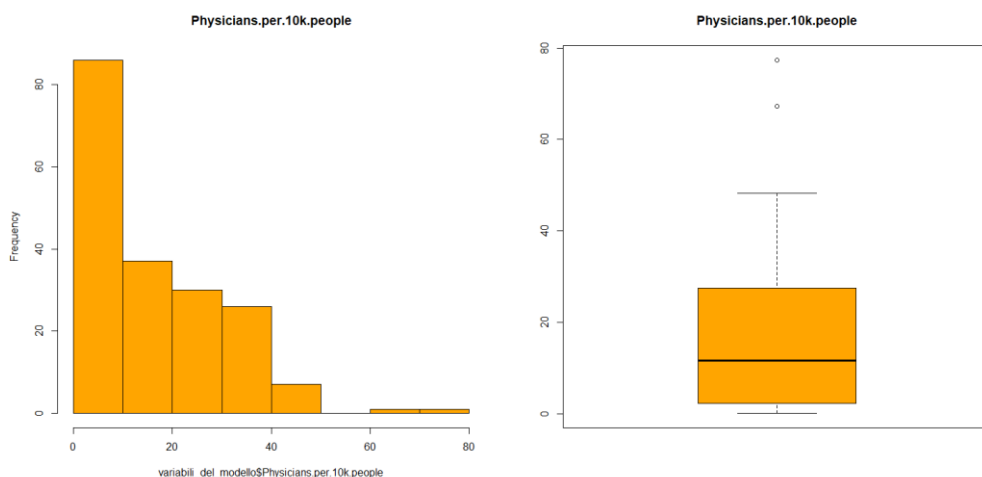


Coerentemente con i valori ottenuti nel modello finale, più alto è il valore GII, maggiori sono le disparità tra donne e uomini e conseguentemente maggiore sarà la diminuzione dell'indice sviluppo umano. Infatti l'odds dell'indice di sviluppo umano medio in corrispondenza di un punto percentuale in più di GII è $\exp(0.01 \cdot -0.9068558) = 0.99$ volte l'analogo odds in corrispondenza di un punto percentuale in meno di GII.

- *Physicians.per.10k.people*

La capillarità dei medici di base sul territorio nazionale rappresenta certamente il livello di qualità del sistema sanitario.

```
> summary(variabili_del_modello$Physicians.per.10k.people)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.080   2.328  11.630  15.688  27.420  77.390
```

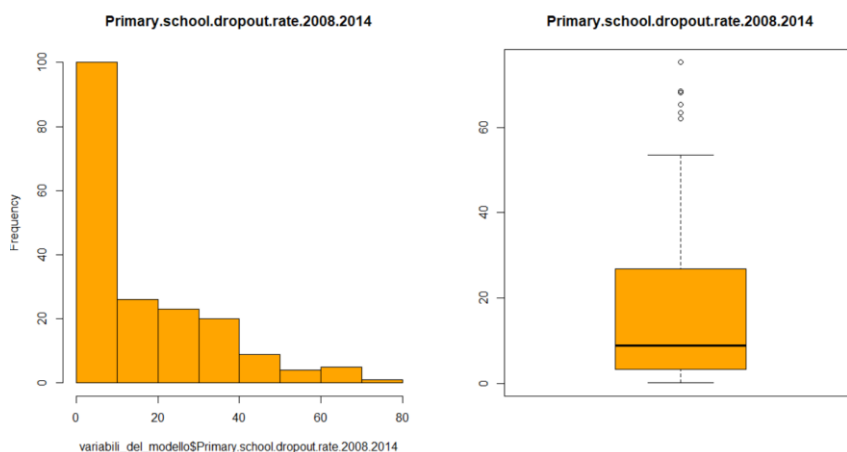


E' preoccupante notare una disparità così ingente tra i due estremi della distribuzione, in particolare tra Tanzania e Qatar che presentano rispettivamente 0.08 e 77.39 medici ogni 10k abitanti.

- *Primary.school.dropout.rate.2008.2014*

Il tasso di abbandono scolastico, è una statistica molto informativa circa le condizioni di un paese dal momento che un elevato tasso di abbandono scolastico potrebbe implicitamente indicare la necessità dei giovani di abbandonare prematuramente gli studi per entrare nel mondo del lavoro, e parallelamente, una scarsa propensione all'innovazione da parte del paese in questione.

```
> summary(variabili_del_modello$Primary.school.dropout.rate.2008.2014)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.1617  3.4196  8.9754 16.2788 26.7172 75.1622
```

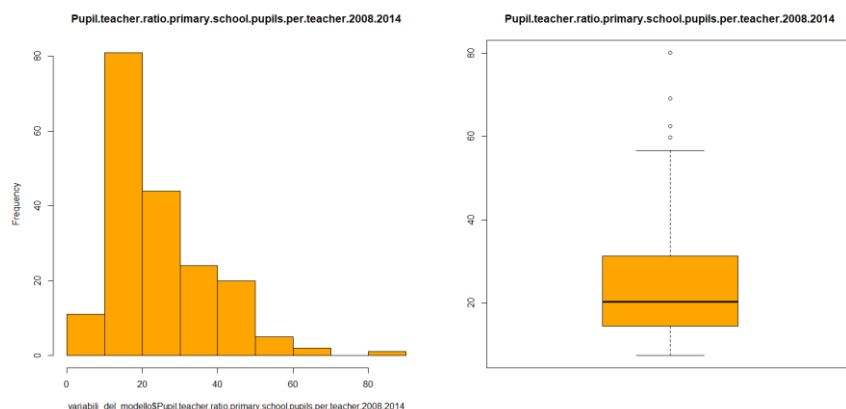


Si osserva che il 25 % dei paesi in esame, ha un tasso di abbandono scolastico superiore al 26.7% fino ad un massimo di 75.16 % registrato in Uganda.

- *Pupil.teacher.ratio.primary.school.pupils.per.teacher.2008.2014*

Il seguente ratio indica la qualità del sistema educativo nazionale, dal momento che un'elevato numero medio di alunni per insegnante è il risultato di risorse carenti a disposizione delle scuole.

```
> summary(variabili_del_modello$Pupil.teacher.ratio.primary.school.pupils.per.teacher.2008.2014)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 7.444 14.493 20.342 24.290 31.360 80.116
```

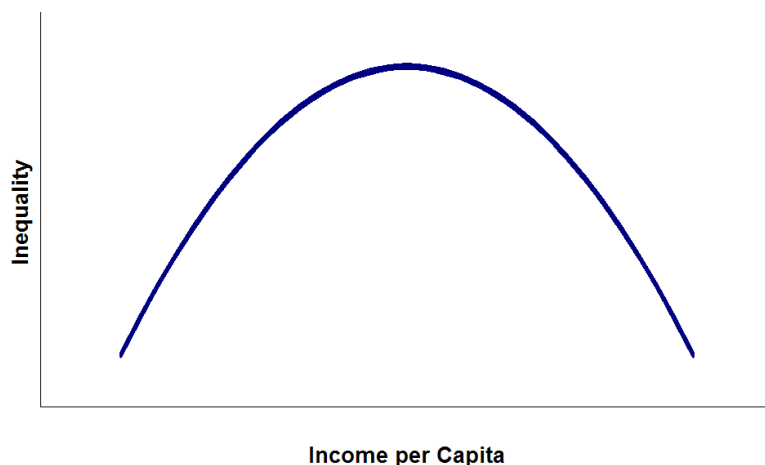


Il 75% della distribuzione presenta un numero medio di alunni inferiore a 32.

CONCLUSIONI E CONSIDERAZIONI DI CARATTERE ECONOMICO

Basandosi sulla nota relazione messa in evidenza dalla **curva di Kuznets** tra disuguaglianza e reddito pro capite nella transizione ad un'economia di tipo industriale, si decide di reinserire l'indice di Gini nel modello dal momento che il reddito è una delle 3 componenti che tramite media geometrica determinano il valore dell'ISU.

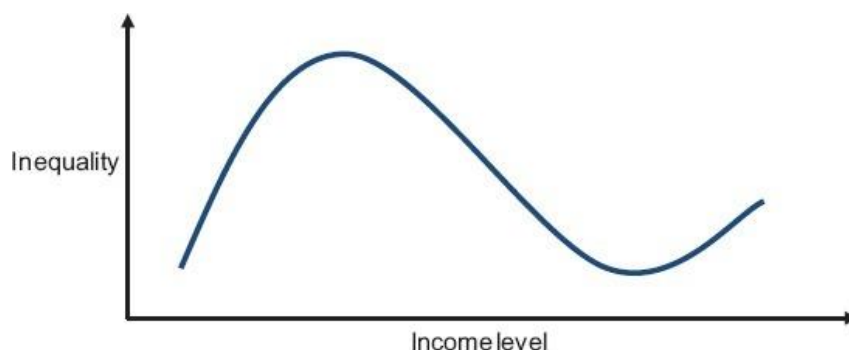
La curva di Kuznets è descritta dal grafico avente sull'asse delle ascisse il Prodotto Nazionale Lordo pro capite e su quello delle ordinate il coefficiente di Gini.



La curva di Simon Kuznets, la cui forma assomiglia ad una U rovesciata, mostra come la distribuzione del reddito tende a peggiorare nella prima fase dello sviluppo, migliorando invece in maniera costante con la transizione a un'economia di tipo industriale. Questo avviene in quanto in una prima fase la fascia di popolazione più ricca investe il proprio capitale, incrementando ulteriormente la propria ricchezza; in un secondo momento però viene colpita maggiormente dalla tassazione, con conseguente effetto redistributivo e miglioramento generale delle condizioni della comunità.

Tuttavia secondo **Piketty**, la tendenza osservata da Kuznets nei primi anni '50 non è necessariamente il risultato di forze economiche intrinseche (ad es. effetti del progresso tecnologico). Infatti, è stato il valore degli immobili anziché la disparità salariale a diminuire negli anni, e ciò è accaduto per ragioni non meramente economiche (ad esempio, la creazione di imposte sul reddito). Di conseguenza la riduzione della concentrazione dei redditi non proseguirebbe necessariamente.

Un esempio significativo riguarda gli Stati Uniti in cui negli ultimi trenta anni le disuguaglianze sono aumentate drasticamente.



Viste le considerazioni di cui sopra, si mostra l'output derivante dall'inserimento della covariata Gini Index nel modello:

```
Call:
betareg(formula = Human.Development.Index.HDI.2014 ~ Electrification.rate.or.population + Internet.users.percentage.of.population.2014 +
+Research.and.development.expenditure..2005.2012 + Physicians.per.10k.people + Gender.Inequality.Index.2014 +
+Primary.school.dropout.rate.2008.2014 + Pupil.teacher.ratio.primary.school.pupils.per.teacher.2008.2014 +
+Gini.coefficient.2005.2013 | Physicians.per.10k.people + Primary.school.dropout.rate.2008.2014 +
+Gender.Inequality.Index.2014 + Pupil.teacher.ratio.primary.school.pupils.per.teacher.2008.2014,
data = selezione_variabili, hessian = T)

Standardized weighted residuals 2:
      Min       IQ      Median       3Q       Max
-2.4565 -0.7207 -0.0320  0.7685  2.7369

Coefficients (mean model with logit link):

              Estimate Std. Error z value Pr(>|z|)
(Intercept)    0.2856545   0.1409945    2.026   0.0428 *
Electrification.rate.or.population 0.0045993   0.0008880    5.179 2.23e-07 ***
Internet.users.percentage.of.population.2014 0.0100709   0.0008013   12.569 < 2e-16 ***
Research.and.development.expenditure..2005.2012 0.1466254   0.0327538    4.477 7.58e-06 ***
Physicians.per.10k.people 0.0044577   0.0007466    5.970 2.37e-09 ***
Gender.Inequality.Index.2014 -0.8909334   0.1113379   -8.002 1.22e-15 ***
Primary.school.dropout.rate.2008.2014 -0.0024748   0.0012932   -1.914   0.0557 .
Pupil.teacher.ratio.primary.school.pupils.per.teacher.2008.2014 -0.0044520   0.0020131   -2.211   0.0270 *
Gini.coefficient.2005.2013 0.3112288   0.1645352    1.892   0.0585 .

Phi coefficients (precision model with log link):

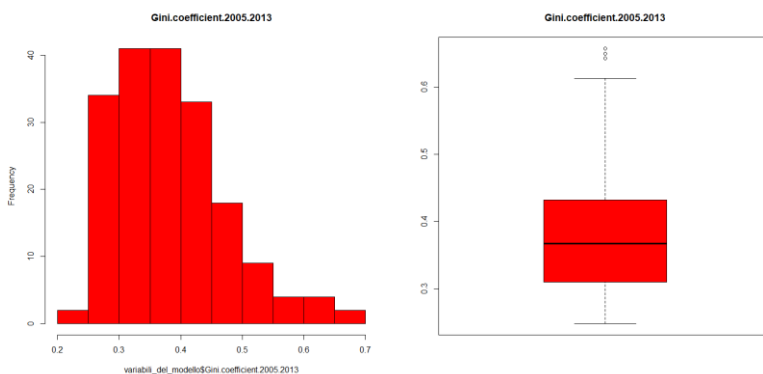
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    2.47013   0.64462    3.832 0.000127 ***
Physicians.per.10k.people 0.06533   0.01495    4.368 1.25e-05 ***
Primary.school.dropout.rate.2008.2014 0.04156   0.01189    3.495 0.000474 ***
Gender.Inequality.Index.2014 3.85328   1.07750    3.576 0.000349 ***
Pupil.teacher.ratio.primary.school.pupils.per.teacher.2008.2014 -0.02988   0.01683   -1.776 0.075764 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Type of estimator: ML (maximum likelihood)
Log-likelihood: 351.8 on 14 Df
Pseudo R-squared: 0.9186
Number of iterations in BFGS optimization: 27
```

L'indice di Gini risulta significativo ad un livello di significatività del 10%.

Tuttavia l'odds dell'indice di sviluppo umano medio in corrispondenza di un punto percentuale in più di GINI INDEX è $\exp(0.01 \cdot 0.3112288) \approx 1$ volte l'analogo odds in corrispondenza di un punto percentuale in meno di GINI INDEX. In altri termini, non sembrerebbe avere alcun effetto sulla determinazione dell'Indice di Sviluppo Umano.

```
> summary(variabili_del_modello$Gini.coefficient.2005.2013)
      Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
 0.2482  0.3101  0.3675  0.3829  0.4322  0.6577
```



Distribuzione Gini.coefficient.2005.2013 nei dati osservati.

Infatti in corrispondenza del valore massimo di Indice di Concentrazione dei redditi della distribuzione (0.6577) registrato nelle isole Seychelles ritroviamo un ISU pari a 0.7724594, che è decisamente sopra il valore medio registrato dai paesi facenti parti dello studio (0.6924). Ad avvalorare ulteriormente questa tesi, si analizza il caso degli Stati Uniti dove in corrispondenza di un Indice di Gini sopra la media, ossia pari a 0.4112 si riscontra un'ISU tra i più elevati (0.9149594).

	Human.Development.Index.HDI.2014	Gini.coefficient.2005.2013
United States	0.9149594	0.4112
Seychelles	0.7724594	0.6577

Si opterà dunque per la rimozione dell'indice di Gini dal modello. Questo offre importanti spunti di analisi circa l'attendibilità stessa dell'Indice di sviluppo umano. L'ISU, infatti, tiene conto solo dei risultati medi di un paese in termini di salute, istruzione e reddito, ma non del modo in cui tali risultati sono distribuiti tra la sua popolazione permettendo così di ottenere una stima "corretta" del valore medio di ciascuna dimensione in base al suo livello di omogeneità.

L'ISU, non considera quindi le disparità in termini di sviluppo umano tra la popolazione all'interno dello stesso paese, motivo per cui due paesi con diverso livello di concentrazione di redditi possono comunque avere lo stesso valore ISU.

L'indice di sviluppo umano è stato criticato anche sotto altri aspetti. In particolare è accusato di riflettere un pregiudizio ideologico contro i modelli di sviluppo economico e sociale dei paesi occidentali industrializzati. Ulteriore critica deriva dal non tener conto delle questioni ambientali riflettendo quasi esclusivamente risultati scolastici e sanitari dei singoli Stati. Infine è accusato di avere un punto di vista poco generico, di dare troppo credito alle statistiche nazionali e di usare metri di giudizio troppo severi e condizionati dalle abitudini di pochi Paesi di riferimento.

Una critica autorevole arriva da *Amartya Sen*, secondo cui nella costruzione dell'Indice di Sviluppo Umano vengono omessi fattori di sviluppo molto importanti quali la presenza di elezioni libere, una stampa indipendente, un sistema politico multipartitico e delle garanzie di libertà di espressione.